5 FAH-5 H-300
DATA ADMINISTRATION

5 FAH-5 H-310
DATA MANAGEMENT SERVICES

(TL:ITS-1:   02-13-2002)

5 FAH-5 H-311 GENERAL

(TL:ITS-1:   02-13-2002)

a. The Data Administration (DA) program is identified by 5 FAM 600 as providing policy, program direction, and standards for Department-wide data to be used in Information Technology (IT) development, integration, and modification projects. This resource management function for the Department's investment in data helps ensure compliance with industry's best practices while maintaining an oversight role on existing systems. The program office, IRM/OPS/SIO/API/DA, may be contacted by phone at (703) 875-4400. Additional information about data administration is also available from the OpenNet.

b. The data administration program fulfills this role with a number of activities grouped under three primary functions: service, standardization and supply. Data administration works hand-in-hand with development and integration activities to provide guidance on data management and data standards. It uses the knowledge gained in that service function to support data standardization, based on the actual business use of data in the Department. The program also serves as a coordinator of data sources and a provider of authoritative data.

c. Administrative costs of the data administration program (including maintenance of the enterprise data model, evaluation of the conclusions of the program through the Data Administration Working Group (DAWG), and technical support to the metadatabase integrated tool set) are funded through the program offices. Fund citations are included in the project-funding request.

5 FAH-5 H-312 PROJECT SERVICES

(TL:ITS-1: 02-13-2002)

The data administration program works hand-in-hand with development and integration initiatives to provide data management expertise in several areas. This provides immediate and continuing benefit to the Department in accomplishing specific goals. It also provides credibility to the program's standards and policies, as they all emerge directly from actual data use in the Department.

5 FAH-5 Table H-312(1)Process Modeling

ACTIVITY	PURPOSE
Process Modeling	a. Process modeling is an analysis tool supporting requirement identification. Facilitated conversations with employees provide answers to questions critical to the understanding of the environment and the purpose for the projected system. The product is the process model—a graphical image of the business process—a diagram indicating the start, the steps and the completion of the activity. This diagram serves as a focus for discussion as the process is validated. The diagram can also identify bottlenecks in the process, repeated steps and other inefficiencies. b. This analysis leads to a set of statements that articulate a desired future state—things that need to change to improve the business process of the office. The answers to these questions point the way to the process, data and other requirements for the new system. c. One of the ways projects go wrong is for the answer to be provided before the question has been asked. In 5 FAM 600, it states that requirements are clearly, unambiguously identified before acquisition and/or development begins. Such effort ensures that the problems are identified before solutions are advanced.

5 FAH-5 Table H-312(2) Data Modeling

ACTIVITY	PURPOSE
Data Modeling	a. Computer systems manipulate data. Ultimately, everything a computer does can be reduced to the motion of switches—binary digits, or "bits"—that are turned "on" or "off." Millions of these switches combine to effect extremely complicated activities—word-processing, spreadsheets, or on-line transactions. For these activities to be effective, the data used must be organized for efficient use. b. The standard industry practice for database organization is normalization. This technique, illustrated below, attempts to ensure that a single piece of information is stored in one—and only one—place, and that information relationships are accurately and unambiguously represented.

For a simple example of data modeling, consider the following typed list of telephone numbers:

5 FAH-5 Table H-312(3) Non-Normalized Data

Name	Type	Number
Charlie Brown	Home	(555) 555-1212
Charlie Brown	Cell	(505) 444-1212
Linus van Pelt	Home	(555) 555-2222
Linus van Pelt	Cell	(505) 444-7474
Linus van Pelt	FAX	(555) 555-2223
Lucy van Pelt	Home	(555) 555-2222

NOTE: There are only three names on the list, and three different types of telephone numbers. Normalizing this list might create a name table, a phone number category table, and a phone number table as shown in 5 FAH-5 Table H-312(4).

5 FAH-5 Table H-312(4) Normalized Data

Name Table	Phone Category Table	Phone Number Table

ID #	Name	ID #	Type	Name	Type	Number
1	Charlie Brown	1	Home	1	1	(555) 555-1212
2	Linus van Pelt	2	Cell	1	2	(505) 444-1212
3	Lucy van Pelt	3	FAX	2	1	(555) 555-2222
		4	Office	2	2	(505) 444-7474
				2	3	(555) 555-2223
				3	1	(555) 555-2222

a. Normalization could further occur by separating first names from surnames, or segmenting phone number area codes (and even by phone exchanges), or isolating only unique phone numbers. These new attributes would add additional fields of information about the data elements.

b. The continuing challenge of normalization is to organize data in ways meaningful to the user while avoiding any repetition of information. One result is that the data can be retrieved, displayed and printed in different ways. Another result is that the computer can manage the relationships between data elements more readily when the data is normalized. Different people have different telephone numbers, establishing one relationship. Different people have different categories of phones and thus, the table above on the right bridges two separate relationships. A third result is the reduction in data redundancy—storage of the same data in more than one place. Along with this is the reduction in data inconsistency. What frequently happens when the same data is stored in several places is that the data values are different, leading to confusion as to which value is correct.

c. In some hardware and/or software environments, optimized data retrieval might require that data be organized in ways specific to the environment. This is known as de-normalization. If a project has made a decision to de-normalize data, the decision and its justification should be documented for future reference.

d. Data modeling identifies data with data names; it describes data with data attributes; and it identifies relationships among data objects, usually referred to as entities. An entity is the item about which you are gathering data. This graphical depiction of data also identifies data cardinality—the quantitative relationship between items; every item A is related to zero, one, or many occurrences of item B; item B may exist independently of item A. Again, using the table above, each person apparently can have as many as four different phone numbers.

e. Graphical data models provide an authoritative map to the information being managed by a system, answering questions and reducing ambiguity. Graphical data models are also a great deal easier to understand than a textual representation of the same information.

f. A data model is a necessary tool for an analyst to understand the overall requirements for a business process. Each step of a business process "handles" data. Data is retrieved, stored, manipulated, and passed on to another part of the process. For the process to operate efficiently, the supporting data must be available and structured to accommodate the process. A data model is a visual way to describe the required data structure.

g. Further, effective normalization and accurately recording data management decisions provide flexibility for the system. A system can be built for one-and-only-one purpose, then very often a need arises for that information to be effectively moved to another environment. Such integration is facilitated by effective data modeling.

5 FAH-5 Table H-312(5) Data Mapping and Integration

ACTIVITY	PURPOSE
Data Mapping and Integration	a. An organization rarely has the opportunity to build everything at once. It is almost inevitable that data mapping will be necessary to combine information from two or more systems, in support of system integration. b. Data mapping involves clearly understanding the data in both systems, and then articulating the way in which the data can be transferred between the systems.

To illustrate the issues associated with data mapping, consider the two data tables in 5 FAH-5 Table H-312(6).
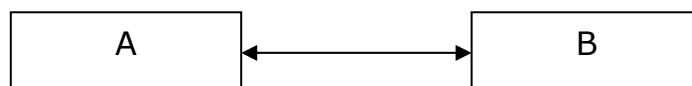
5 FAH-5 Table H-312(6) Data Mapping

System A		System B			
Name	**Phone**	**LName**	**FName**	**Area Code**	**Number**
Linus Van Pelt	(555) 555-2222	Bailey	Beetle	444	444-8686
Lucy Van Pelt	(555) 555-2222	Bumstead	Blondie	777	707-3030
Charlie Brown	(555) 555-1212	Bumstead	Dagwood	777	707-3030

h. Data mapping between these two systems will involve constructing several procedures, known as algorithms, for moving the data from system to system. If information is going to move from system A into system B, system A's "Name" field will have to be broken into first and surnames; likewise, system A's "Phone" field will have to be separated

into the "Area Code" and "Number" entities in system B. The opposite
procedures would be required for migrating information out of system B
to system A. The construction of these data tables and algorithms is the
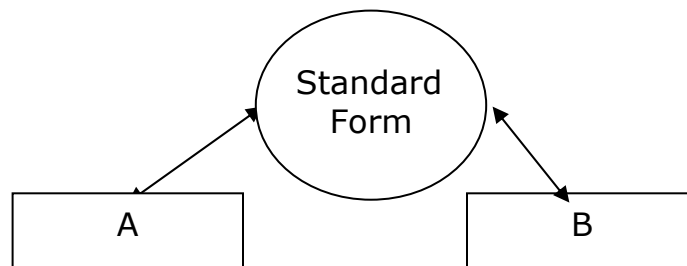data mapping process.

i. It is important to recognize, as well, that a direct map of system to
system is not considered industry best practice because it permanently
links the two systems together in ways that may be counter-productive.
The physical constraints of the two systems become such that no changes
can be made to either system without the changes impacting them both
simultaneously. 5 FAH-5 Table H-312(7) illustrates this "direct linkage."

5 FAH-5 Table H-312(7) Hard-wired System Integration

```
┌──────────────┐                    ┌──────────────┐
│      A       │◄──────────────────►│      B       │
└──────────────┘                    └──────────────┘
```

j. By establishing the data standard as the integration point, each system
needs only to continue to maintain the data map between the system and
the standard. The result is conceptually shown in 5 FAH-5 Table H-
312(8).

5 FAH-5 Table H-312(8) Mapping Through the Data Standard

```
              ┌─────────────┐
              │  Standard   │
              │    Form     │
              └─────────────┘
             ▲               ▲
            /                 \
┌──────────────┐          ┌──────────────┐
│      A       │          │      B       │
└──────────────┘          └──────────────┘
```

k. Using a standard form of name for data objects, along with a standard
form for the data contained in data objects, reduces the complexity of the
algorithm and makes data mapping easier. Thus, the "standardization"
process described below makes data management easier. For guidance
refer to the *Object Definition and Naming Standard*, available from the
program office or on the Data Administration OpenNet site at
http://da.irm.state.gov/dataadministration/publications.asp.

l. Commercial off-the-shelf products create unique problems in data
mapping. Because commercial products are designed to address a
specific and finite series of functions, rather than to fit comfortably within
a suite of software systems, additional analysis is necessary to enable the

integration. Industry best practice typically requires that process and data models of the commercial product be delivered along with the product itself. Where such documentation is unavailable, it becomes necessary to study the product at length to generate the background for data mapping to occur. The total cost of a commercial off-the-shelf product can be raised significantly by the analysis, modeling and mapping work required to effectively integrate it into the enterprise.

5 FAH-5 Table H-312(9) Data Quality Analysis

ACTIVITY	PURPOSE
Data Quality Analysis	a. A business is not merely a collection of processes, it is also a collection of business rules—business policies that govern its own behavior and distinguish it from others. These rules govern changes in the state of the enterprise, and apply specifically to data elements. When business rules are not clearly articulated, the user community implies them—different users may therefore imply different things, leading to misunderstandings and error. Data quality, then, is interpreted in consideration of how consistent data is with the business rules of the enterprise. b. Data quality audits can identify the extent to which a database is consistent with its own business rules, but does not automatically solve the problems involved e.g., knowing that a business rule exists that every customer address must contain a ZIP code does not provide ZIP codes for the 43% of the addresses missing them. In many cases, enterprises must accept databases audited to internal consistency below 50% because the time, expense and sheer ability to correct the problems are not available.

5 FAH-5 H-313 DATA STANDARDIZATION

(TL:ITS-1: 02-13-2002)

a. If a program was intended to perform the "service" functions described above, it would make a significant contribution to an enterprise. However, this contribution would be limited if it were not tied together in

meaningful ways. The data administration program, therefore, uses its
"service" component as the data gathering mechanism for
standardization. By working with actual data use in the Department, data
administrators better understand the data objects, attributes,
relationships, cardinality, and business rules of the Department. By
conducting further analysis, data administration organizes related
information and articulates this actual business usage as the standard for
the Department. This provides guidance for new systems and integration
activity. The standardization effort also maintains flexibility in the
enterprise for data reuse and elimination of data redundancy.

5 FAH-5 Table H-313(1) Enterprise Data Model

ACTIVITY	PURPOSE
The Enterprise Data Model (and Standard Data Elements)	a. Information gathered about data usage in the Department moves into the enterprise data model. A continuing work in progress, the enterprise data model is regularly updated in quarterly releases of the *Standard Data Elements* volume [available from the Data Administration OpenNet site at http://da.irm.state.gov/dataadministration/publications.asp. This document provides data models graphically depicting data objects and their relationships, and articulates standardized data names, data attributes and business rules relevant to the data objects. b. The enterprise data model is **not** intended as a requirement, but as a statement of how data is used in the Department. An office wishing to use the model for employee names would probably not use all of the elements in the "person name" model, which explains all the data requirements identified within the Department. c. The guidance of the data administration program is that all development and integration activity use the standard data elements articulated in the enterprise data model whenever possible, and especially for data integration as shown above. Where questions emerge about how to apply the

| | enterprise data model in a particular environment, contact the data administration program. |

5 FAH-5 Table H-313(2) MetaDataBase

ACTIVITY	PURPOSE
MetaDataBase	a. The metadatabase is the integrated set of data tools used by the data administration program to store the information contained in the enterprise data model. Data models, process models, relational databases and other forms provide a comprehensive view of data usage in the Department. b. System developers may wish to use the metadatabase repository as a common source of information for system development. This topic is discussed in the *Repository Implementation Guidelines* document published by Data Administration and available on the OpenNet site at http://da.irm.state.gov/dataadministration/publications.asp.

5 FAH-5 Table H-313(3) Data Administration Standardization

ACTIVITY	PURPOSE
Data Administration Working Group (DAWG)	a. In order to ensure that data administration's information about data usage in the Department is generalized beyond one specific office environment, the Data Administration Working Group meets quarterly to discuss additions suggested to the "standard data elements" document—candidate standard data elements— as well as other topics of common interest. In these sessions, recommended data names and data formats are viewed in the context of other business users, so that the resulting standard can be generally beneficial.

	b. Meetings of the Data Administration Working Group are open to all who are interested in attending. Database administrators and data stewards are particularly encouraged to attend.
	c. Information or proposals may be submitted to the Data Administration Working Group by contacting the data administration program. Questions about the Data Administration Working Group should likewise be directed to the data administration program at (703) 875-4400.

5 FAH-5 H-314 SUPPLY

(TL:ITS-1: 02-13-2002)

Data administration is the resource management function for Department data usage. As the only program studying data throughout the Department, data administration is uniquely positioned to identify opportunities for data re-use. The "supply" function is the third major component of the data administration program.

5 FAH-5 Table H-314(1) Standard Data Tables

ACTIVITY	PURPOSE
Standard Data Reference Tables	In many cases, data has a single source and changes little over time. In such cases, data administration makes an effort to manage the data and provide it to the enterprise in useful form. The Data Administration OpenNet site contains databases of several types of reference data for general use by users and developers throughout the Department.

5 FAH-5 Table H-314(2) Data Stewardship

ACTIVITY	PURPOSE
Data Stewardship	a. More commonly, the members of a specific business area manage data. The Bureau of Human Resources manages information about

the employee and the Bureau of Financial Management and Policy manages financial information. In such cases, it is unnecessary for data administration to take ownership of the information. Those in the specific business area become the data stewards, providing access to this information for Departmental use.

b. Data stewards identify the conditions whereby business users in certain roles should be allowed to create, read, update, and/or delete information. They also manage the data quality of the database. Data stewards facilitate data re-use, and move the Department closer to the goal of reducing data redundancy while supporting integration.

5 FAH-5 H-315 THROUGH H-319 UNASSIGNED